

⑩ 日本国特許庁(JP)

⑪ 特許出願公開

⑫ 公開特許公報(A)

昭64-88875

⑬ Int. Cl.⁴

G 06 F 15/38
3/16

識別記号

3 4 0

庁内整理番号

V-7313-5B
Q-7341-5B

⑭ 公開 昭和64年(1989)4月3日

審査請求 未請求 発明の数 1 (全5頁)

⑮ 発明の名称 音声翻訳装置

⑯ 特 願 昭62-246149

⑰ 出 願 昭62(1987)9月30日

⑱ 発 明 者 松 浦 博 神奈川県川崎市幸区柳町70番地 株式会社東芝柳町工場内
⑱ 発 明 者 渡 辺 貞 一 神奈川県川崎市幸区柳町70番地 株式会社東芝柳町工場内
⑲ 出 願 人 株 式 会 社 東 芝 神奈川県川崎市幸区堀川町72番地
⑳ 代 理 人 弁 理 士 鈴 江 武 彦 外2名

明 細 書

1. 発明の名称

音 声 翻 訳 装 置

2. 特許請求の範囲

(1) 入力音声を認識する音声認識部と、この音声認識された言語情報を他国語の言語情報に翻訳する翻訳部とを具備した音声翻訳装置において、

翻訳前の言語情報と翻訳後の言語情報とをその出力メディアを異ならせて併せて出力することを特徴とする自動翻訳装置。

(2) 出力メディア機器は音声出力装置と表示装置とからなり、翻訳前後の言語情報を音声および表示文字列としてそれぞれ併せて出力するものである特許請求の範囲第1項記載の音声翻訳装置。

3. 発明の詳細な説明

[発明の目的]

(産業上の利用分野)

本発明は異なる言語音声間での対話を実現する音声翻訳装置に関する。

(従来の技術)

使用言語の異なる当事者間で、相互にその使用言語を用いながら対話することは人類の長年の夢である。このような対話を実現するものとして音声翻訳装置が考えられている。この音声翻訳装置は実用化が進められている音声認識装置や機械翻訳装置、音声合成装置等を用いて実現されるもので、例えば日本語・英語間での対話を実現する場合には次のように構成される。

即ち、音声入力された日本語を日本語音声認識部にて認識し、この認識された日本語言語情報を日英翻訳部にて英語言語情報に翻訳する。そしてこの英語言語情報を英語音声合成部にて英語音声に合成変換し、これを相手側に音声出力する。一方、相手側から音声入力される英語を英語音声認識部にて認識し、この認識された英語言語情報を英日翻訳部にて日本語言語情報に翻訳する。そしてこの日本語言語情報を日本語音声合成部にて日本語音声に合成変換し、これを音声出力するように構成される。

このような音声翻訳装置を用いることにより、一方の利用者は日本語を音声入力しながら相手側からの通話情報を日本語音声として聞き、また他方の利用者は英語を音声入力しながら相手側からの通話情報を英語音声として聞くことが可能となり、ここにその翻訳通信(対話)が実現される。

ところがこのような音声翻訳装置を実際に実現し、これを運用するに際しては様々な問題が生じる。その最も大きな課題は入力音声に対する認識とその翻訳処理であり、複数の認識(翻訳)候補や誤認識(翻訳)が生じることが多々ある。この結果、翻訳されて音声出力される言語情報の言語的ニュアンスが翻訳前の言語音声を持つニュアンスと異なってしまう弊れが多分にあった。

このような言語ニュアンスの異なる翻訳音声が出力されると、その対話において誤解が生じたり、不快感が生じることがあり、折角の音声翻訳による対話の効果が損われると云う不具合が生じる。また翻訳ニュアンスの異なりから、翻訳出力される言語音声だけではその対話意図を十分理解する

ことができないと云う問題も生じた。

(発明が解決しようとする問題点)

このように従来にあっては音声翻訳装置を介して異種言語音声により対話しようとする場合、翻訳出力される言語音声は翻訳前の言語音声を持つ言語的ニュアンスを必ずしも正確に反映していないと云う問題があり、無用な誤解や不快感を生じさせる虞れがあった。

本発明はこのような事情を考慮してなされたもので、その目的とするところは、誤解や不快感を生じさせることなく、また相手側の対話意図を十分に把握しながら異種言語音声間での対話を効果的に進めることを可能とする音声翻訳装置を提供することにある。

[発明の構成]

(問題点を解決するための手段)

本発明は入力音声を認識する音声認識部と、この音声認識された言語情報を他国語の言語情報に翻訳する翻訳部とを具備した音声翻訳装置において、

翻訳前の言語情報と翻訳後の言語情報とをその出力メディアを異ならせて、例えば音声出力装置と表示装置とを用いて上記翻訳前後の言語情報を音声および表示文字列としてそれぞれ併せて出力することを特徴とするのである。

(作用)

本発明によれば、翻訳前後の言語情報が、例えば文字列および音声等として異なるメディアにより併せて出力されるので、例えば音声出力された言語情報を聞きながら文字列として表示出力された言語情報を視認することができる。つまり視覚と聴覚とを併用して翻訳前の言語情報と翻訳後の言語情報とを対比することが可能となる。この結果、利用者(対話者)は異種言語に対する多少の知識を持ち合わせるだけで、相手側の言語的ニュアンスを把握することが可能と、無用な誤解や不快感の発生を効果的に防ぐことが可能となる。

(実施例)

以下、図面を参照して本発明の一実施例につき説明する。

第1図および第2図は本発明の実施例に係る音声翻訳装置の基本的な構成例を示す図であり、例えば電話端末間での音声翻訳通信に供される装置として実現される。ここで電話端末1の利用者の使用言語が日本語であり、電話端末2の利用者の使用言語が英語であって、音声翻訳装置は日本語・英語間での音声翻訳を行なうものとする、音声翻訳装置は次のように構成される。

日本語音声認識部3は電話端末1から音声入力される日本語音声を音声認識するもので、日英翻訳部4はこの認識された日本語言語情報を英語言語情報に翻訳する。また英語音声認識部5は電話端末2から音声入力される英語音声を音声認識するもので、英日翻訳部6はこの認識された英語言語情報を日本語言語情報に翻訳する。

第1図に示す音声翻訳装置では、上述した如く翻訳された言語情報を英語音声合成部7にて英語音声に合成変換して電話端末2に音声出力し、また日本語音声合成部8にて日本語音声に合成変換して電話端末1に音声出力するものとなっている。

これに加えて電話端末1側では、電話端末1に備えられた表示部9を用いて前記英語音声認識部5で求められた日本語翻訳前の言語情報である英語文を、前述した日本語音声出力と合せて表示するものとなっている。そして電話端末2側では、電話端末2に備えられた表示部10を用いて前記日本語音声認識部3で求められた英語翻訳前の言語情報である日本語文を、前述した英語音声出力と合せて表示するものとなっている。

つまり電話端末1側では音声翻訳された日本語が音声出力されると共に、その翻訳前の英語文が文字列表示されるようになっており、他方の電話端末2側では音声翻訳された英語が音声出力されると共に、その翻訳前の日本語文が文字列表示されるようになっている。

これに対して第2図に示す如く構成された音声
翻訳装置にあっては、電話端末1,2からそれぞれ
入力される言語音声そのまま相手側の電話端末
2,1に対して音声出力される。そして電話端末1
に備えられた表示部9にて前記英口翻訳部8で翻

て回線に送出される。この初期通信モードによってその電話端末の構成や通信しようとする情報の態様（直接音声の通信か翻訳通信か）、翻訳の形態（翻訳言語の指定）等の通知がなされ、また通信回線の接続制御が行われる。このとき、必要なメッセージ情報等は前記ディスプレイ18を介して表示出力される。

さてマイクロフォン15を介して入力された音声はA/D変換器16を介して取込まれ、データメモリ17に格納されると共に、音声分析部18にてフィルタリング等の音響分析が施される。セグメント変換部19は標準パターンメモリ20を参照して前記音響分析結果から、例えば音素や音節、またはVCV単位の音声認識の為のセグメント情報を求めている。音声認識部21はこのセグメント情報に従い、認識辞書22を参照して前述した入力音声を認識処理している。この音声認識処理は、DPマッチングや遷移ネットワーク等を用いて行われる。この際、必要に応じて音声の再入力が促される。このようにして求められた認識結果(言語情報)

訳されて求められた言語情報である日本語文を文字列表示し、電話端末2に歸えられた表示部10にて前記日英翻訳部4で翻訳されて求められた言語情報である英語文を文字列表示するものとなっている。

つまり電話端末1側では音声入力された翻訳前の英語がそのまま音声出力されると共に、英口翻訳部6で求められた翻訳後の日本語文が文字列表示されるようになっており、他方の電話端末2側では音声入力された翻訳前の日本語がそのまま音声出力されると共に、日英翻訳部4で求められたその翻訳後の英語文が文字列表示されるようになっている。

このようなシステム構成を採用して実現される
音声翻訳機能付き電話端末は、例えば第3図に示
すように構成される。

第3図において、11は制御部、12はキー入力部、13はディスプレイである。音声翻訳通信に先立ち、キー入力部12から所定のキー入力が行なわれると、その入力情報は制御部11から網終端装置14を介し

は、例えば文節単位毎に区分される等して前記データメモリ17に適宜格納される。

翻訳部23は翻訳辞書24を参照して上述した如く認識された言語情報を翻訳処理するものである。この翻訳処理は、例えば日英翻訳や英日翻訳等、予め定められた言語間での翻訳のみならず、この翻訳通信システムにおいて共通に設定された中間言語との間での翻訳を行なう場合もあるが、一般的にはその翻訳処理の形態は翻訳通信端末毎に設定される。このようにして翻訳処理された言語情報が前記網終端装置14を介して通信回線に送出される。このとき上記翻訳処理に供された言語情報（翻訳前の情報）も前記網終端装置14を介して通信回線に送出される。

一方、通信回線から網終端装置14を介して受信される言語情報に対して制御部11は翻訳前の言語情報と翻訳後の言語情報とを前記データメモリ17に別々に格納する。規則合成部25は規則合成辞書26を参照して、例えば翻訳後の言語情報に対する音韻・韻律パラメータ系列を生成している。音声

合成部27はこのような音韻・韻律パラメータ系列に従って音声信号を規則合成により生成し、D/A変換器28を介して出力している。このようにして規則合成された音声信号によってスピーカ29が駆動されて合成音声が発せられることになる。また翻訳前の言語情報は、そのまま文字列情報として前記ディスプレイ13に表示されるようになってゐる。

このような一連の処理は、プログラムメモリ80に格納された制御プログラムに従い、前記制御部11の制御の下で実行される。

尚、ここでは翻訳後の言語情報を音声合成して出力し、翻訳前の言語情報を文字列表示するものとするが、前述したように翻訳前の言語情報を音声合成して出力し、翻訳後の言語情報を文字列表示することも可能である。

このように電話端末を構成することによって、例えば日本語入力された音声は英語情報に翻訳され、その翻訳前後の言語情報が他方の翻訳通信端末に通信される。そして相手側の電話端末にて翻

訳前後の言語情報が音声および文字列として、そのメディアを異ならせて合せて出力される。また他方の電話端末から英語で音声入力された情報は日本語情報に翻訳され、その翻訳前後の言語情報がそれぞれ通信回線に送出される。そしてその翻訳前後の言語情報が音声および文字列として、そのメディアを異ならせて合せて出力される。この結果、音声言語情報と文字列言語情報とを併用し、これらのメディアの違いを利用して翻訳前後の言語情報をそれぞれ併せ出力して、日本語と英語との間の音声翻訳通信が行われる。

尚、電話端末がA/D変換器18とD/A変換器28とからなる音声コーデックだけを備えて構成され、この音声コーデックを網終端装置14を介して通信回線に接続して構成される場合には、例えば第5図に示すように、その通信回線上に中央翻訳システムを設ける。そしてこの中央翻訳システムにて上述した翻訳処理を行なわせるようにし、この中央翻訳システムを中継して前記電話端末に翻訳前後の言語情報をそれぞれ与えるようにすれば

良い。

尚、この中央翻訳システムは、前述した音声分析部18、セグメント変換部19、標準パターンメモリ20、音声認識部21、認識辞書22、翻訳部23、翻訳辞書24、規則合成部25、規則合成辞書26、音声合成部27、そしてデータメモリ17とプログラムメモリ30を備えて構成されるものである。

このようにして翻訳前後の言語情報を音声および文字列としてそれぞれ併せて出力する本装置によれば、音声出力される言語情報を聞きながら文字列表示される言語情報を見て、その翻訳前後の言語情報を対比することが可能となる。

従って対話相手が使用する言語に対する多少の知識を備えておれば、翻訳出力された言語情報だけでは汲取ることの困難な言語的ニュアンスを容易に把握することが可能となる。この結果、無用な誤解が生じることを防ぎつつ、対話相手のニュアンス(意図)を理解しながら異種言語間での対話を進めることが可能となる。

尚、本発明は上述した実施例に限定されるもの

ではない。例えば文字列の表示手段としてはプリンタ等のハードコピー装置であっても良い。また例示した言語以外の言語に対する翻訳を行なうものであっても良い。

更には入力音声の認識処理や翻訳処理の方式、また音声合成の方式については従来より種々提唱されている方式をシステム仕様に応じて採用すれば良いものである。その他、本発明はその要旨を逸脱しない範囲で種々変形して実施することができる。

【発明の効果】

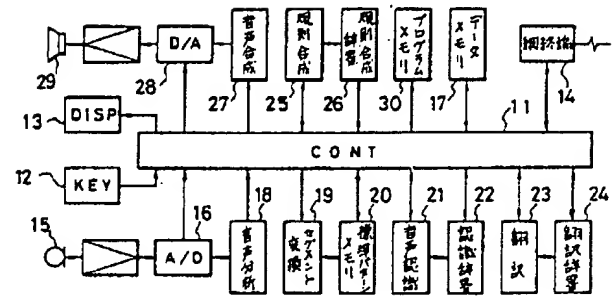
以上説明したように本発明によれば、翻訳前後の言語情報を音声と文字列としてそのメディアを異ならせて併せ出力するので、翻訳前後の言語情報を効果的に対比することが可能となる。この翻訳前後の言語情報の対比によって翻訳ニュアンスの違いが生じるような場合であっても相手側の意図を容易に汲取ることが可能となるので、異種言語間での対話を誤解を生じることなしに効果的に進めることが可能となる等の実用上多大なる効

果が奏せられる。

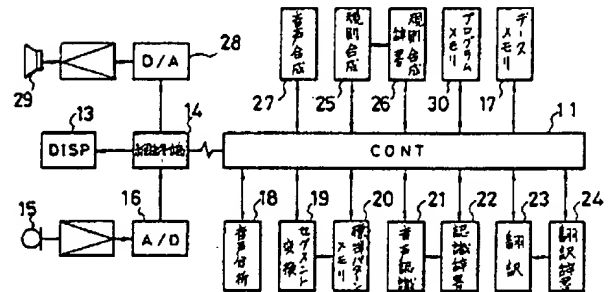
4. 図面の簡単な説明

第1図および第2図はそれぞれ本発明の実施例に係る音声翻訳装置の基本的な構成例を示す図、第3図は音声翻訳機能を備えた電話端末の構成例を示す図、第4図は本装置を実現する他の例を示す図である。

1…電話端末、3…日本語音声認識部、4…日英翻訳部、5…英語音声認識部、6…英日翻訳部、7…日本語音声合成部、8…英語音声合成部、9,10…表示部、13…ディスプレイ、14…網終端装置、15…マイクロホン、16…A/D変換器、18…音声分析部、19…セグメント変換部、21…音声認識部、23…翻訳部、25…規則合成部、27…音声合成部、28…D/A変換器、29…スピーカ。

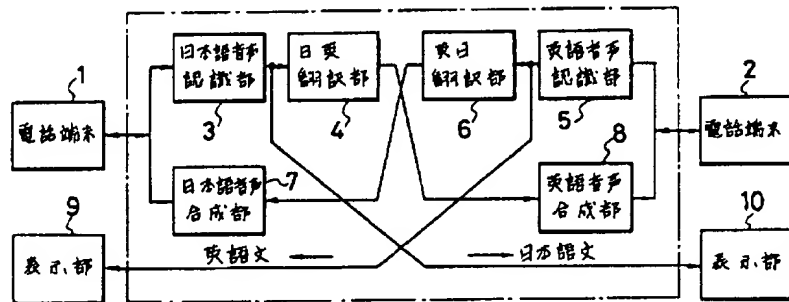


第3図

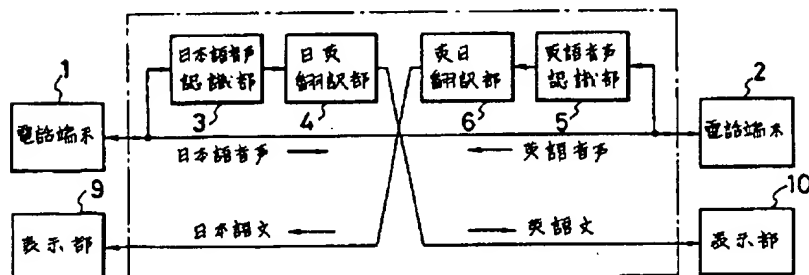


第4図

出願人代理人 弁理士 鈴江武彦



第1図



第2図

JP-S64-88875

A patent application public release S64-88875
Public April 3, 1989

Title of the invention

Speech translation device

Inventor: Hiroshi Matsuura, Teiichi Watanabe
Applicant TOSHIBA

Example

It is explained about one embodiment of the invention when taken with the drawing as follows.

Figure 1,2 are basic configuration example of speech translation device hanging in example of the present invention.

By way of example only, it is realized as device to be able to leave by speech translation communication at a telephone handset interval.

The language which user of telephone handset 1 employs here is Japanese.

The language which a user of telephone handset 2 employs is English.

Speech translation device does speech translation at Japanese / English interval.

Speech translation device is configured as follows to do these.

Japanese speech recognizer 3, speech recognition does a voice input Japanese voice to be able to go away from telephone handset 1.

Japanese-to-English compilation department 4 translates this recognized Japanese linguistic information into English linguistic information.

In addition, English speech recognizer 5, speech recognition does a

voice input English voice to be able to go away from telephone handset 2.

English-to-Japanese compilation department 6 translates this recognized English linguistic information into Japanese linguistic information.

In speech translation device shown in figure 1, synthesis converts translated linguistic information to English voice with English voice synthesis department 7 as depicted above.

And voice output is done to telephone handset 2.

In addition, synthesis is converted to a Japanese voice with Japanese voice synthesis department 8.

And voice output is done to telephone handset 1.

Display part 9 comprised to telephone handset 1 is used in telephone handset 1 side as well as this.

"The English statement that is linguistic information before Japanese translation pursued in above mentioned English speech recognizer 5" is displayed in total with previously described Japanese voice output. And, in telephone handset 2 side, display part 10 comprised to telephone handset 2 is used.

The Japanese statement that is linguistic information before English translation pursued in said article Japanese speech recognizer 3 is displayed in total with previously described English voice output. In other words, in telephone handset 1 side, the Japanese which was able to leave speech recognition can leave voice output. In addition, it is displayed a character string an English statement before the translation. In other telephone handset 2 side, the English that was able to leave speech translation can leave voice output. In addition, it is displayed a character string a Japanese statement before the translation. As against this, speech translation device shown in figure 2 is explained in the following.

The speech sound voice that it is input from telephone handset 1,2 respectively can just leave voice output as against telephone handset 2,1 of a counterpart.

And, in display part 9 comprised to telephone handset 1, it is displayed a character string the Japanese statement that is the linguistic information which it is translated with above-mentioned English and Japanese translation part 6, and was demanded.

In representation 10 comprised to telephone handset 2, a character string displays the English statement that is the linguistic information which it is translated with above-mentioned Japan and Britain translation part 4, and was demanded.

In other words English before the translation which was able to leave voice input can just leave voice input in telephone handset 1 side. Along with it, it is displayed a character string a Japanese statement after translation pursued with English-to-Japanese compilation department 6.

In other telephone handset 2 side, Japanese before the translation which was able to leave voice input can just leave voice output.

Along with it, it is displayed a character string English statement after the translation pursued with Japanese-to-English compilation department 4.

The speech translation facility that such a system configuration is adopted, and is realized is gained, and, for example, telephone handset is configured as shown in figure 3.

In figure 3, there are 11 control section, 12 key input unit and 13 displays. Speech translation communication is preceded, a prescribed key input is done from key input unit 12.

The output information goes through network termination 14 from control section 11, and it is emitted by line.

By means of this initial communicate mode, it seems to become the following. "Configuration of telephone handset and status (direct sound vocal communication or translation communication) of the information which is going to communicate", configuration (a specification of translation language) of translation are notified of.

In addition, connection control of communication line is done.

Message information necessary then goes through above

mentioned D play 13, and it is provided an output to indicate.

By the way, the voice that mic 15 is gone through, and it was input goes through A/D converter 16, and it is taken. It is stored to data memory 17. In addition, an acoustic analysis such as filtering is given with speech analysis department 18.

Segment converter 19 refers to standard pattern 1 memory 20.

And "music, a syllable or segment information for speech recognition of a VCV unit" is demanded from above mentioned acoustic analysis result. Speech recognizer 21 acts on on in this segment information.

It recognizes, and a previously described input sound voice is handled referring to recognition dictionary 22.

This speech recognition process is done by means of dynamic programming matching method or a succession network.

In doing so, as necessary phonic reentry is promoted.

In this way, for example, demanded recognition result (linguistic information) is distinguished every clause unit.

It is stored to above mentioned data memory 17 appropriately by these processes. Compilation department 23 refers to translation dictionary 24. Linguistic information recognized as above is translated.

This translation is not only Japanese-to-English translation and English-to-Japanese compilation (translation at a predetermined language interval).

In this translation communication system, translation at an interval with an intermediate language set by commonness is done.

Configuration of the translation is generally set by every translation communications terminal.

In this way translated linguistic information goes through above-mentioned network termination 14, and it is emitted by communication line.

Linguistic information (information before translation) offered for the translation then goes through network termination 14, and it is

emitted by communication line.

On the other hand, the linguistic information which network termination 14 is gone through, and is received is processed as follows by communication line.

Control section 11 stores "linguistic information before translation" and "linguistic information after translation" to above mentioned data memory 17 separately.

Speech synthesis by rule part 25 refers to speech synthesis by rule dictionary 26.

It generates a phoneme / a prosody parameter list as opposed to linguistic information after translation.

Voice synthesis department 27 acts on on in a subsidiary of such phoneme / prosody parameter, and audio signal is generated by speech synthesis by rule.

And it is output through D/A converter 28.

In this way speaker 29 is driven with the audio signal which was able to leave speech synthesis by rule.

And it is expected that synthesized speech is emitted.

In addition, linguistic information before translation is just displayed to above mentioned display 13 as subsidiary of character information.

Such a serial process acts on a stored control program to program memory 30.

It is carried out with control of above mentioned control section 11.

Voice synthesis does linguistic information after translation here, and it is output.

And it is assumed that a character string displays linguistic information before translation.

However, voice synthesis does linguistic information before translation as-previously-described, and it is output.

And a character string can display linguistic information after translation.

As thus described the voice that was able to leave Japanese input is

translated into English information by arranging telephone handset.

And, the translation anteroposterior linguistic information, it is communicated by the other translation communications terminal.

And translation anteroposterior linguistic information makes the media be different as a voice and a character string in telephone handset of a counterpart.

And it is put together, and it is output.

In addition, the information that was able to leave voice input in English is translated from the other telephone handset into Japanese information.

The translation anteroposterior linguistic information is emitted by communication line respectively.

And the translation anteroposterior linguistic information makes the media be different as a voice and a character string, and it is output in total.

As a result, character string linguistic information is used together with spoken language information.

A difference of these medias is used.

And translation anteroposterior language glossary is put together respectively, and it is output.

At Japanese and an English interval, speech translation communication is done.

Telephone handset comprises only A/D converter 16 and voice Coe Dick comprising 28 D/A converter,, and it is configured.

Network termination 14 is gone through, and this voice Coe Dick is connected to communication line.

For this case, for example, being similar provide central translation system on the communication line as shown in figure 5.

And it makes do above mentioned translation in this central translation system.

And this central translation system is relayed, linguistic information of translation front and back should be given above mentioned telephone handset respectively.

This central translation system is configured by the following.
Previously described speech analysis department 18
Segment converter 19 , Standard pattern memory 20 , Department of voice vocabulary 21, Lexical dictionary 22 , Compilation department 23, Translation dictionary 24 , Speech synthesis by rule part 25 , Speech synthesis by rule dictionary 26 , Voice synthesis department 27, Data memory 17, Program memory 30.
These are possessed, and it is configured.

In this way, according to this device which, at the same time, output translation anteroposterior linguistic information as a voice and a character string respectively, the following can say.

While hearing the linguistic information which can leave voice output, it is language Get Info displayed a character string.

And the translation anteroposterior linguistic information gets possible to be compared.

Thus, if some factual knowledge as opposed to the language which an interaction partner employs is comprised, a nuance of difficult language of what translation grasps only from output linguistic information gets possible to be grasped easily.

As a result, useless misunderstanding is prevented from occurring.

And while understanding a nuance of an interaction partner, interaction at a heterologous language interval gets possible to be pushed forward.

The present invention is not a thing limited to above mentioned example.

By way of example only, it may be hardcopy device such as a printer for displaying means of a character string.

In addition, compilation as opposed to language aside from the language which illustrated may be done.

About "an input sound vocal recognition process, a method of translation, a method of a voice synthesis", it is processed as follows.

A system specification is accepted, and a method proposed in these

system in various ways conventionally should be adopted.

In addition, the present invention transforms it in the range that does not deviate from the subject matter, and it can be embodied.

Brief description of drawings

Figure 1 , 2

Basic configuration example of speech translation device hanging in example of the present invention

Figure 3

Configuration example of telephone handset comprising a speech translation facility

Figure 4

Other examples to realize this device.

- (1) (2) telephone handset
- (3) a Japanese speech recognizer
- (4) Japanese-to-English compilation department
- (5) an English speech recognizer
- (6) English-to-Japanese compilation department
- (7) Japanese voice synthesis department
- (8) English voice synthesis department
- (9) (10) display part
- (13) display
- (14) network termination
- (15) microphone
- (16) A/D converter
- (18) speech analysis department
- (19) segment converter
- (21) speech recognizer
- (23) compilation department

- (25) speech synthesis by rule part
- (27) voice synthesis department
- (28) D/A conversion department
- (29) speaker